International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST218415

Hybrid Heart Disease Prediction Supervised, Unsupervised Opinion Mining Algorithms

Miss. Samiksha Arvind Kale*, Prof. Dr A .B .Gadicha
Department of Computer Science and Engineering, P. R. Pote (Patil) College of Engineering & Management, Amravati, Maharashtra, India

Article Info
Volume 8, Issue 4
Page Number : 36-41

Publication Issue
July-August-2021

Article History
Accepted : 01 July 2021
Published : 05 July 2021

ABSTRACT

Heart plays significant role in living organisms. Diagnosis and prediction of heart related diseases requires more precision, perfection and correctness because slightly mistake can cause fatigue problem or death of the person, there are numerous death cases related to heart and their counting is increasing exponentially day by day. To affect the matter there's essential need of prediction system for awareness about diseases Machine learning is that the branch of AI (AI), it provides prestigious support in predicting any quite event which take training from natural events. During this paper, we calculate accuracy of machine learning algorithms for predicting heart condition, for this algorithms are k-nearest neighbor, decision tree, linear regression and support vector machine (SVM) by using UCI repository dataset for training and testing. For implementation of Python programming Anaconda (jupytor) notebook is best tool, which have many kind of library, header file, that make the work more accurate and precise.

Keywords : Heart Disease, Data Mining, Classification, Supervised, Unsupervised, Linear regression, decision tree

I. INTRODUCTION

Heart is one of the foremost extensive and vitals of physical body therefore the care of heart is vital. Most of diseases are related to heart therefore the prediction about heart diseases is important and for this purpose comparative study needed during this field, today most of patient are died because their diseases are recognized eventually stage because of lack of accuracy of instrument so there's need to realize the more efficient algorithms for diseases prediction. Machine Learning is one of the efficient technology for the testing, which is based on training and testing. It is the branch of Artificial Intelligence (AI) which is one of broad area of earning where machines emulating human abilities, machine learning could also be a selected branch of AI. On the other hand machines learning systems are trained to seek out out the thanks to process and make use of data hence the mixture of both technology is additionally called as Machine Intelligence. because the definition of machine learning, it learns from the

phenomenon, natural things so during this project we uses the biological parameter as testing data like cholesterol, vital sign, sex, age, etc. and on the thought of these, comparison is completed within the terms of accuracy of algorithms like during this project we've used four algorithms which are decision tree, linear regression, k-neighbour, SVM. During this paper, we calculate the accuracy of 4 different machine learning approaches and on the thought of calculation we conclude that which one is best among them.

Section 1 of this paper consist the introduction about the machine learning and heart diseases. Section II described, the machine learning classification. Section III illustrated the related work of researchers. Section IV is about the methodology used for this prediction system. Section V is about the algorithms utilized during this project. Section VI briefly describes the dataset and their analysis with the results of this project. And thus the last Section VII concludes the summary of this paper with slight view about future scope of this paper.

II. RELATED WORK

Heart is one of the core organ of human body , it play crucial role on blood pumping in human body which is as essential because the oxygen for human body so there's always need of protection of it, this is often one of the massive reasons for the researchers to work on this. So there are number of researchers performing on it .There is always need of study of heart related things either diagnosis or prediction otherwise you'll say that protection of heart disease .There are various fields like AI , machine learning, processing that contributed on this work .

Performance of any algorithms depends on variance and biasness of dataset [4]. As per research on the machine learning for prediction of heart diseases himanshu et al. [4] naive bayes perform well with low variance and high biasness as compare to high variance and low biasness which is knn. With low biasness and high variance knn suffers from the matter of over fitting this is often the rationale why performance of knn get decreased. There are various advantage of using low variance and high biasness because the dataset small it take less time for training also as testing algorithm but there also some disadvantages of using small size of dataset. When the dataset size get increasing the asymptotic errors are get introduced and low biasness, low variance based algorithms play well during this type of cases. Decision tree is one of the nonparametric machine learning algorithm but as we all realize it suffers from the matter over fitting but it cloud be solve by some over fitting removable techniques. Support vector machine is algebraic and statics background algorithm, it construct a linear separable n-dimensional hyper plan for the classification of datasets.

The nature of heart is complex, there's need of carefully handling of it otherwise it cause death of the person. The severity of heart diseases is assessed supported various methods like knn, decision tree, generic algorithm and naïve bayes [3]. Mohan et al. [3] define how you'll combine two different approaches to make one approach called hybrid approach which have the accuracy 88.4% which is sort of all other.

Some of the researchers have worked on processing for the prediction of heart diseases. Kaur et al. [6] have worked on this and define how the interesting pattern and knowledge are derived from the huge dataset. They perform accuracy comparison on various machine learning and processing approaches for locating which one is best among then and acquire the result on the favor of svm.

Kumar et al.[5] have worked on various machine learning and processing algorithms and analysis of these algorithms are trained by UCI machine learning dataset which have 303 samples with 14 input feature

and located svm is best among them, here other different algorithms are naivy bayes, knn and decision tree.

Gavhane et al. [1] have worked on the multilayer perceptron model for the prediction of heart diseases face to face and therefore the accuracy of the algorithm using CAD technology. If the amount of person using the prediction system for his or her diseases prediction then the notice about the diseases is additionally going to increases and it make reduction within the death rate of heart patient.

Some researchers have work on one or two algorithm for predication diseases. Krishnan et al. [2] proved that decision tree is more accurate as compare to the naïve bayes classification algorithm in their project. Machine learning algorithms are used for various kind of diseases predication and much of the researchers have work on this like Kohali et al.[7] work on heart diseases prediction using logistic regression, diabetes prediction using support vector machine, carcinoma prediction using Adaboot classifier and concluded that the logistic regression give the accuracy of 87.1%, support vector machine give the accuracy of 85.71%, Adaboot classifier give the accuracy up to 98.57% which good for predication point of view.

A survey paper on heart diseases predication have proven that the old machine learning algorithms doesn't perform good accuracy for the predication while hybridization perform good and provides better accuracy for the predication[8]..

III. METHODOLOGY OF SYSTEM

Processing of system start with the data collection for this we uses the UCI repository dataset which is well verified by number of researchers and authority of the UCI [15].

A. Data Collection

First step for predication system is data collection and deciding about the training and testing dataset. During this project we've used 73% training dataset and 37% dataset used as testing dataset the system.

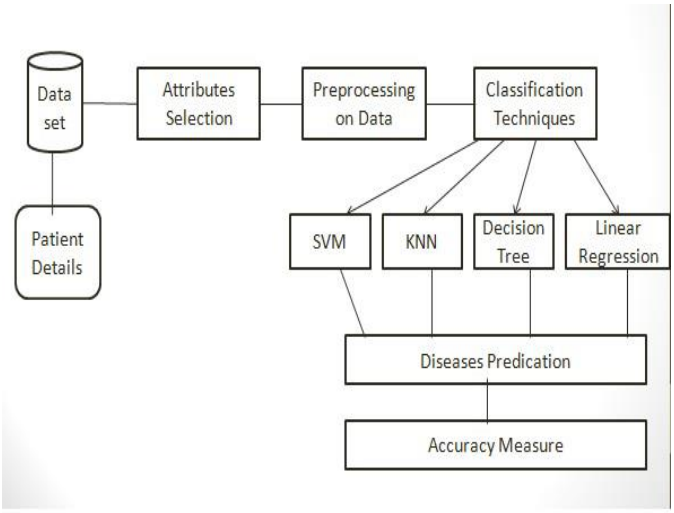

Figure.1 Architecture of Prediction System

B. Attribute Selection

Attribute of dataset are property of dataset which are used for system and for heart many attributes are like heart bit rate of person, gender of the person, age of the person and much of more shown in TABLE.1 for predication system.

C. Pre-processing of data

Pre-processing needed for achieving prestigious result from the machine learning algorithms. as an example Random forest algorithm doesn't support null values dataset and for this we've to manage null values from original data.

For our project we've to convert some categorized value by dummy value means within the type of "0"and "1" by using following code:

D. Data Balancing

Data balancing is vital for accurate result because by data balancing graph we'll see that both the target classes are equal. Fig.3 represents the target classes where "0" represents with heart diseases patient and "1" represents no heart diseases patients.

S. No.	Attribute	Description	Type
1	Age	Patient's age (29 to 77)	Numaric
2	Sex	Gender of patient(male-0 female-1)	Nominal
3	Cp	Chest pain type	Nominal
4	Trestbps	Resting blood pressure(in mm Hg on admission to hospital ,values from 94 to 200)	Numerical
5	Chol	Serum cholesterol in mg/dl, values from 126 to 564)	Numerical
6	Fbs	Fasting blood sugar>120 mg/dl, true-1 false-0)	Nominal
7	Resting	Resting electrocardiographics result (0 to 1)	Nominal
8	Thali	Maximum heart rate achieved(71 to 202)	Numerical
9	Exang	Exercise included agina(1-yes 0-no)	Nominal
10	Oldpeak	ST depression introduced by exercise relative to rest (0 to .2)	Numerical
11	Slope	The slop of the peak exercise ST segment (0 to 1)	Nominal
12	Ca	Number of major vessels (0-3)	Numerical
13	Thal	3-normal	Nominal
14	Targets	1 or 0	Nominal

TABLE.1 Attributes of the Dataset

III. RESULT

Accuracy of the algorithms are depends on four values namely true positive (TP), false positive (FP), true negative (TN) and false negative (FN).

$$\text{Accuracy} = \frac{(FN+TP)}{(TP+FP+TN+FN)} \quad (1)$$

The numerical value of TP, FP, TN, and FN defines as:
TP= Number of person with heart diseases
TN= Number of person with heart diseases and no heart diseases
FP= Number of person with no heart diseases
FN= Number of person with no heart diseases and with heart diseases

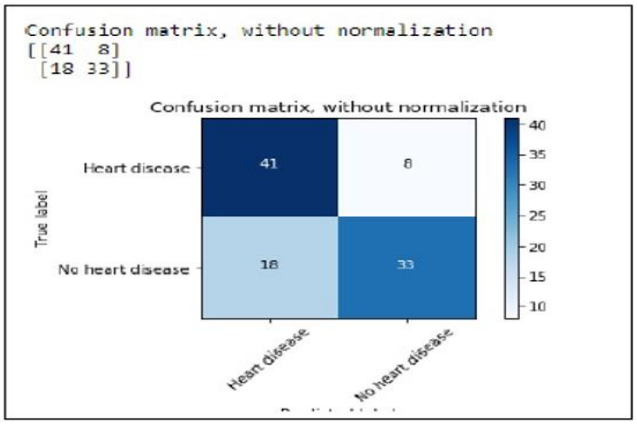

Figure.2 Confusion matrix for Decision tree

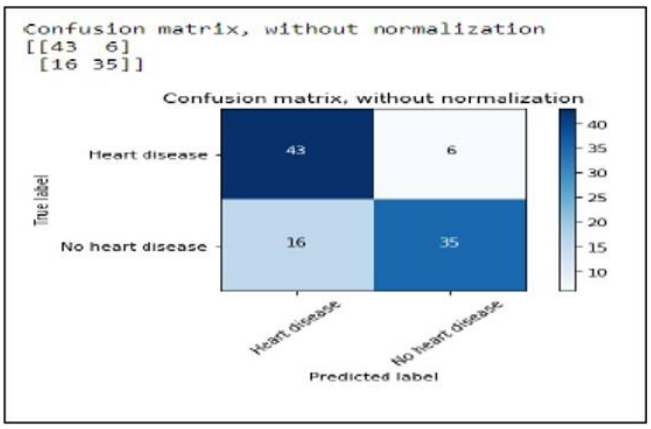

Figure.3 Confusion Matrix for linear regression

After performing the machine learning approach for testing and training we find that accuracy of the knn is much efficient as compare to other algorithms. Accuracy should be calculated with the support of

confusion matrix of each algorithms here number of count of TP, TN, FP, FN are given and using the equation (1) of accuracy, value has been calculated and it is conclude that knn is best among them with 87% accuracy and the comparison is shown in TABLE.2

Algorithm Accuracy	Algorithm Accuracy
Support Vector machine 83%	Support Vector machine 83%
Decision tree 79%	Decision tree 79%
Linear regression 78%	Linear regression 78%
k-nearest neighbor 87%	k-nearest neighbor 87%

Table.4.3 Accuracy comparison

IV. CONCLUSION

Heart is one of the essential and vitals of human body and prediction about heart diseases is additionally important concern for the citizenry so as that the accuracy for algorithm is one of parameter for analysis of performance of algorithms. Accuracy of the algorithms in machine learning depends upon the dataset that used for training and testing purpose. Once we perform the analysis of algorithms on the thought of dataset whose attributes are shown in TABLE.1 and on the thought of confusion matrix, we discover KNN is best one.

For the long run Scope more machine learning approach are getting to be used for best analysis of the centre diseases and for earlier prediction of diseases so as that the speed of the death cases are often minimized by the notice about the diseases.

V. REFERENCES

[1] Santhana Krishnan J and Geetha S, "Prediction of Heart Disease using Machine Learning Algorithms" ICIICT, 2019.

[2] Aditi Gavhane, Gouthami Kokkula, Isha Panday, Prof. Kailash Devadkar, "Prediction of Heart Disease using Machine Learning", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology(ICECA), 2018.

[3] Senthil kumar mohan, chandrasegar thirumalai and Gautam Srivastva, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" IEEE Access 2019.

[4] Himanshu Sharma and M A Rizvi, "Prediction of Heart Disease using Machine Learning Algorithms: A Survey" International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 8 , IJRITCC August 2017.

[5] M. Nikhil Kumar, K. V. S. Koushik, K. Deepak, "Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools" International Journal of Scientific Research in Computer Science, Engineering and Information Technology ,IJSRCSEIT 2019.

[6] Amandeep Kaur and Jyoti Arora,"Heart Diseases Prediction using Data Mining Techniques: A survey" International Journal of Advanced Research in Computer Science , IJARCS 2015-2019.

[7] Pahulpreet Singh Kohli and Shriya Arora, "Application of Machine Learning in Diseases Prediction", 4th International Conference on Computing Communication And Automation (ICCCA), 2018.

[8] M. Akhil, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm," Procedia Technol., vol. 10, pp. 85–94, 2013.

[9] S. Kumra, R. Saxena, and S. Mehta, "An Extensive Review on Swarm Robotics," pp. 140–145, 2009.

[10] Hazra, A., Mandal, S., Gupta, A. and Mukherjee, " A Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review" Advances in Computational Sciences and Technology , 2017.

[11] Patel, J., Upadhyay, P. and Patel, "Heart Disease Prediction Using Machine learning and Data Mining Technique" Journals of Computer Science & Electronics , 2016.

[12] Chavan Patil, A.B. and Sonawane, P."To Predict Heart Disease Risk and Medications Using Data Mining Techniques with an IoT Based Monitoring System for Post-Operative Heart Disease Patients" International Journal on Emerging Trends in Technology, 2017.

[13] V. Kirubha and S. M. Priya, "Survey on Data Mining Algorithms in Disease Prediction," vol. 38, no. 3, pp. 124–128, 2016.

[14] M. A. Jabbar, P. Chandra, and B. L. Deekshatulu, "Prediction of risk score for heart disease using associative classification and hybrid feature subset selection," Int. Conf. Intell. Syst. Des. Appl. ISDA, pp. 628–634, 2012.

[15] https://archive.ics.uci.edu/ml/datasets/Heart+Disease

Cite this article as :

Miss. Samiksha Arvind Kale, Prof. Dr A .B . Gadicha, "Hybrid Heart Disease Prediction Supervised, Unsupervised Opinion Mining Algorithms", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 8 Issue 4, pp. 36-41, July-August 2021. Available at doi : https://doi.org/10.32628/IJSRST218415
Journal URL : https://ijsrst.com/IJSRST218415